

The AP Course Audit Syllabus Review Process

Methodological Explanation

© 2009 by the Educational Policy Improvement Center. ALL RIGHTS RESERVED.
Any use of materials on this website, including reproduction, modification, distribution or republication, without the prior written consent of EPIC, is strictly prohibited.



April
2007

David T. Conley, Ph.D.
*Chief Executive Officer,
Educational Policy Improvement Center*

*Professor,
University of Oregon*

Educational Policy Improvement Center (EPIC)

Dr. David T. Conley, Chief Executive Officer
720 East 13th Street, Suite 202
Eugene, OR 97401
Phone 541.346.6153 • Toll-free 877.766.2279
<http://www.epiconline.org>

Design Description and Rationale

The AP Course Audit utilizes a criterion-based professional judgment method of analysis within a nested multi-step review process. The overall goal of the methodology is to yield a final judgment on each syllabus that is ultimately valid. While reviewer consistency is an important consideration, the most important goal is to reach a final judgment on each syllabus that is accurate. The process is specifically designed to pay special attention to false negatives while simultaneously controlling for false positives.[§]

This general approach allows complex documents, such as course syllabi, to be reviewed in order to reach a holistic, dichotomous judgment based on scores on individual criteria. In this case, the syllabus must meet all curricular requirements in order to serve as the basis for AP course authorization. This approach allows reviewers to utilize their expert judgment, which derives from their deep knowledge of the subject area being reviewed, but avoids pitfalls inherent in holistic judgment approaches to document review, in which a reviewer reaches a decision on the document as a whole. The criterion-based process utilizes rules that constrain reviewer judgment by filtering all decisions through the screen of a series of criteria that are derived directly from published AP Curricular Requirements. To inform reviewer judgment further, decision rules accompany each subject and most curricular requirements. The final result is an overall determination that the syllabus meets standard or does not meet standard.

The use of a criterion-based professional judgment model allows a teacher to fulfill the curricular requirements by submitting one document, a syllabus, rather than being required to fill out questionnaires or inventories where they would be expected to account for how they addressed each curricular requirement. While a standardized reporting method such as a questionnaire or inventory lends itself to greater consistency of scoring, it sacrifices the holistic and integrated view of the AP course being reviewed. The result of fragmenting the review process would be that the syllabi themselves would not necessarily be much improved. One of the goals of the AP Course Audit, in addition to verifying the alignment of the syllabi with the curricular requirements, is to help teachers gain useful information on how to modify the course syllabus so that it better reflects the relevant curricular requirements for the course. Therefore, a review of actual syllabi confirms degree of alignment with the requirements and identifies specific areas where additional information may be needed to demonstrate alignment.

[§] A false negative occurs when a syllabus is designated as not meeting the standard when in fact it actually meets the standard. A false positive occurs when a syllabus is designated as meeting the standard when in fact it does not actually meet the standard.

Achieving these goals requires considerable interpretation by trained reviewers, meaning that judgment is an integral part of the review process.

The AP Course Audit does in fact employ a checklist that does not require any reviewer judgment; this is the AP Course Audit Form. On this form, the teacher indicates adherence to all relevant curricular requirements. If the intent were to avoid reviewer judgment altogether, the AP Course Audit Form would suffice as the most reliable form of reporting possible. However, the form sacrifices validity because it does not allow the reviewer to verify teacher assertions or to judge individual assertions in context.

In short, the use of the criterion-based professional judgment method allows for the review of complex documents that can be submitted intact without the need for the teacher to provide additional evidence. This method verifies teacher assertions provided on the AP Course Audit Form or identifies areas where additional information is needed. In the process teachers either have their syllabi validated as reflecting curricular requirements or receive guidance on where they need to add information to demonstrate alignment with the published curricular requirements.

How the Review Process Addresses Issues of Validity and Reviewer Consistency

The review process employs multiple techniques and processes to help ensure that the ultimate decision reached on each individual syllabus is accurate and consistent. This is accomplished through two primary means: 1) a series of procedures to help maximize reviewer consistency, and 2) a nested review process by which each course's syllabus can be submitted up to three times and reviewed by up to seven independent reviewers (the most highly trained reviewers are used in the later cycles), with opportunities for corrective action at the end of each review cycle.

The procedures undertaken to maximize reviewer consistency include the following:

- ◆ Decision rules are developed by senior reviewers in consultation with College Board and methodological consultants. Senior reviewers are selected based on demonstrated experience and expertise with the specific AP subject they will review.
- ◆ The decision rules are developed by having the senior reviewers review and score actual course syllabi and then discuss their scores for each curricular requirement. The process continues over multiple syllabi until agreement is reached on how each curricular requirement should be scored

in the review process. Their reasoning and rationale are recorded and summarized to formulate the decision rules.

- ◆ As a result of this process, senior reviewers develop common mental models of the curricular requirements in the context of a range of course syllabi. Senior reviewers continue to deepen their understandings of the decision rules and how they are to be applied in specific situations through reviews of syllabi that were submitted in advance. They are able to discuss and clarify points of potential ambiguity. This process sharpens consistency among senior reviewers.
- ◆ Reviewers are recruited who possess the content knowledge in the subject area sufficient to determine whether the syllabus conforms to the curricular requirements and to make judgments and inferences necessary for this determination. They have experience teaching the college equivalent of the AP course in question, and many have prior experience as AP exam readers as well.
- ◆ Each reviewer must pass a rigorous training process before being allowed to review documents. The reviewer first reads and rates two practice syllabi against the course requirements, receiving feedback on each to help improve accuracy of judgments.
- ◆ The reviewer then reviews and rates up to three syllabi until receiving a passing score on two of the three syllabi. For most subject areas, the reviewer can only be discrepant with the scoring key on one item. The reviewer receives further explanation from a senior reviewer on any item that is not scored correctly.
- ◆ Finally, the reviewer confers with a senior reviewer if necessary to develop a full and complete understanding of the decision rules and their application to syllabi in the subject area.
- ◆ After passing the training process, the reviewer begins reviewing syllabi. In order to ensure that reviewers' judgments do not drift from the desired criteria, a series of "benchmark" syllabi are periodically inserted into the queue in each subject area. All reviewers review these benchmark syllabi, which have been pre-scored by senior reviewers. The reviewers do not know when they are looking at a benchmark syllabus. The reviewer must receive a passing score on the benchmark syllabus in order to continue to review syllabi. If a benchmark is not passed, the reviewer engages in a conversation with a senior reviewer and must successfully score another benchmark document before being allowed to review any other syllabi.
- ◆ Every syllabus lacking evidence of one or more curricular requirements is reviewed by a senior reviewer.
- ◆ A final check of reviewer consistency is achieved through the "back-read" process in which every twentieth syllabus that has been authorized in each subject is read by an independent reviewer who is unaware that he or she is reviewing a syllabus that has already been considered authorized

by another reviewer. The reviewers do not know when they are engaging in a back-read. This process helps identify the rate of false positives.

The result of the training is to create in the reviewer's mind a mental model of a syllabus that meets all curricular requirements in the subject area. This mental model enables the reviewer to make appropriate inferences where necessary in cases where the decision rules require complex, multi-step, multi-level demonstrations of evidence.

© 2009 by the Educational Policy Improvement Center. ALL RIGHTS RESERVED.
Any use of materials on this website, including reproduction, modification, distribution or republication, without the prior written consent of EPIC, is strictly prohibited.

Given the complexity of the syllabi and the decision rules, the goal of the process is not necessarily to get 100% consistency on the first review. Some syllabi will fall into a “gray area” on one or more curricular requirements. This is inherent in a process designed to retain teacher discretion to pursue sophisticated learning goals and individualized teaching methods.

The Three-Stage Nested Review Process

The review process in its entirety is designed to ensure that no syllabus that deserves approval is ultimately denied approval. This goal is accomplished through the provision of three review cycles. In each of these cycles, any syllabus deemed not to meet standard in one or more areas is viewed by a minimum of two separate and independent reviewers. At the conclusion of any review cycle in which the syllabus did not provide sufficient information to demonstrate alignment with one or more curricular requirements, the teacher receives a report detailing the area or areas in which additional information is required and the reason why more information is being required.

Therefore, it is possible for a syllabus that does not contain sufficient detail or information, particularly in areas where the curricular requirements and decision rules are complex, to be viewed differently by two independent reviewers on one or more curricular requirements.

Stage 1

Whenever a reviewer identifies any element of a syllabus as not meeting the curricular requirements, that syllabus is automatically re-reviewed by a senior reviewer, who has extensive training in the decision rules used to judge each curricular requirement. The senior reviewer can affirm or overturn the determination of the reviewer that the syllabus did not meet standard in a particular area. The rate of reviewer decisions being overturned by senior reviewers is one measure of reviewer consistency.

Stage 2

After the teacher is informed that more information is required, the teacher makes the necessary changes and resubmits the syllabus, which now goes to a new reviewer for a stage 2 review. This reassignment to a new reviewer helps address the issue of potential severity bias by the original reviewer (and senior reviewer). At this point, if the reviewer again indicates that the syllabus lacks information to demonstrate a curricular requirement, it is again reviewed by a senior reviewer and, if the senior reviewer concurs with the reviewer, the syllabus is reviewed independently by staff in the Advanced Placement offices of the College Board. These staff are all expert in the AP course in the subject area in question as well as in the decision rules for the subject, which they helped formulate. Their role is to be the third independent reviewer of the syllabus. If the AP staff member agrees with the reviewer and senior reviewer, the AP staff member contacts the teacher directly and offers focused, specific feedback and suggestions on how to change the syllabus so that it addresses the curricular requirements and decision rules. This interaction provides an additional opportunity to identify misclassified syllabi if the teacher is able to explain or point out something in the syllabus that the three reviewers did not notice or interpret as the teacher intended. In no case will a teacher be asked to provide additional evidence for a requirement that was not identified as needing such additional evidence after the first review.

Stage 3

The teacher now may submit the syllabus for a third review. Once again, the syllabus is assigned to a new reviewer who does not know that the syllabus has been reviewed twice previously. This helps avoid prejudging the syllabus. The reviewer follows the same criterion-based process to rate the syllabus against the curricular requirements. If the syllabus fails to provide sufficient information to demonstrate all necessary curricular requirements yet again, the syllabus goes to another senior reviewer one final time. If the senior reviewer concurs with the reviewer's judgment, the teacher must wait until the summer of 2007 to resubmit the syllabus. During this period of time, the teacher will be encouraged to attend an AP professional development event to receive help and support in constructing an appropriate syllabus that results in a high quality AP course.

Estimating Error in the Process

In essence, this approach, like most scoring methods, is a probabilistic model. The key question is: What are the probabilities that a syllabus that meets standard will be improperly denied by seven independent reviewers over the course of three separate review cycles? A general estimate can be reached after formulating a series of assumptions about reviewer accuracy. These assumptions derive from data on

reviewer performance. Additionally, the assumptions take into account the training process, benchmark document performance, number of reviews overturned by senior reviewers, and back-scoring consistency rates.

Taking these factors into account, an estimate can be made of the number of syllabi correctly and incorrectly identified at each review cycle and the number of misclassified syllabi at the conclusion of three cycles. The following is an estimate of correct classification and misclassification based on the process described.

Given the nature of the rating process in which a series of independent reviews are conducted, it is possible to estimate the likelihood of misclassifying any particular syllabus at the end of the entire rating process as not meeting the criteria when, in fact, it does meet the standards by using a compound probability model. Simply put, since each rating is independent, the likelihood of repeatedly misclassifying a syllabus that should be approved over a number of ratings is the product of the likelihood of misclassifying the syllabus at each rating. Thus, if the probability of misclassification at each stage is constant at “p”, the likelihood of misclassification over “r” number of ratings is p^r .

The proposed training process requires reviewers to correctly classify, at minimum, 2 of 3 training syllabi, or a misclassification rate of 1 of 3, or 33%. Using this rate as a liberal estimate of errors in classification at each rating event, the probability of misclassifying a syllabus following seven ratings would be $(.33)^7$, or .00043. That is, the likelihood of misclassifying a syllabus all seven times would be about .04%, or less than a tenth of one percent. Thus, about 99.96% of the syllabi would be correctly classified using this process. Applied to 100,000 syllabi, it could be estimated that approximately 42-43 syllabi may be misclassified as not meeting the criteria following the seven ratings.

While these estimates of misclassification probabilities are quite low, they may still be reduced in a number of ways. First, the above estimates assume a constant misclassification rate across all seven ratings. In reality, the process employs senior reviewers at three of the seven rating stages and uses College Board personnel in one of the ratings. Given that these reviewers are more experienced and better trained on the criteria for evaluating the syllabi, their misclassification rates may well be less than the misclassification rates of standard reviewers. With even marginal reductions in these misclassification rates for these reviewers (from .33 to .25), the overall misclassification probabilities drop from .04% to .014%. Stated differently, the expected number of misclassifications for 100,000 syllabi drops from approximately 42-43 to about 14, for a correct classification rate of 99.986%. There may be further reductions in the estimated error rates. If all reviewers’ misclassification rates were estimated to be .25

(rather than .33), the rate of misclassification over all seven ratings would be even lower, at .006%, resulting in an estimated 6 syllabi of 100,000 being misclassified. Our experience with the actual error rate for standard reviewers is .15 for false negatives. Using this as an estimate for all reviewers, the likelihood that a syllabus that should be authorized will repeatedly be misclassified as not meeting the standard over the entire rating process is reduced to .0002%, or less than 1 in 100,000 syllabi.

It is worth noting that other methods for scoring complex pieces of work, such as AP free-response items, utilize different means to ensure consistency that were not available to this process. For AP exams, the scores given by the scorer to the free response items on any individual exam are combined with that individual's score on the multiple-choice portion of the exam, thereby lessening the effect of any scorer inconsistency on the ultimate AP Exam grade. The strength of the relationship between scores on the free response items and the multiple-choice items helps establish the final cut points for each level of the exam. In this fashion, reviewer judgment is subsumed within a total score that exists in five performance bands.

Balancing Consistency with Validity

Because the AP Course Audit leads to a dichotomous decision, meets or does not meet, and not a scale score, it is held to a higher standard of consistency. The review process is specifically designed to address consistency issues while accommodating complex documents and expert professional judgment. This allows teachers to accomplish two goals simultaneously with their syllabi: 1) demonstrate that they are meeting the curricular requirements for the course, and 2) retain the flexibility to teach the course in the way they see fit. Any further standardization of the review process would decrease reviewer variance even more, but would also decrease variance among syllabi concomitantly. The process as currently designed and implemented seeks a reasonable balance between these two competing goals—consistency with AP Course Requirements and flexibility to teach the AP course in the manner the teacher finds most appropriate.